

TESTES DE MÚLTIPLA ESCOLHA E DE RESPOSTA LIVRE EM FÍSICA GERAL *

BERNARDO BUCHWEITZ **

RESUMO

São comparados os resultados de questões de múltipla escolha com os de questões de resposta livre de provas aplicadas em um curso de Física Geral. As questões comparadas envolveram diferentes níveis de comportamento e foram elaboradas com base em objetivos previamente definidos. Da análise dos dados obtidos no presente trabalho, concluímos que, essencialmente, não há diferença entre avaliar por múltipla escolha e avaliar por resposta livre.

SUMMARY

Multiple choice and essay tests in Physics. The scores of multiple choice and essay items applied in an introductory college Physics course, are compared. Different behaviours are involved in the test items and its construction is founded on educational objectives previously defined. No difference has been detected in the scores and this result permit us to conclude that there are no significant differences between essay and multiple-choice types of evaluation.

O PROBLEMA

Nas últimas décadas, houve um aumento da população estudantil em todos os níveis de ensino e, com isso, problemas educacionais relacionados com o elevado número de alunos são comuns na atualidade brasileira, inclusive a nível universitário.

Soluções para estes problemas têm sido procuradas através de diversos sistemas de ensino, com tendências ao ensino individualizado, por um lado, e ao ensino de massa, por outro. Ligada a esses diferentes sistemas de ensino e às suas constantes mudanças está a avaliação. Ela pode ser parte integrante de processo educacional — avaliação formativa — ou servir para testar desempenhos e dar graus — avaliação somativa (Block, 1971). Não queremos, no presente trabalho, discutir a necessidade e a importância de avaliar mas, apenas, dentro do amplo e complexo problema, pesquisar em torno do uso de dois tipos de instrumentos de medida

destinados a verificar a ocorrência ou não da aprendizagem. Mais especificamente, pretendemos comparar a avaliação através de questões de múltipla escolha com a avaliação através de questões de resposta livre¹.

A comparação entre esses dois tipos de questões foi feita dentro de três diferentes categorias de comportamento: conhecimento, compreensão e aplicação. Não foram incluídas na investigação questões que exigem comportamentos mais complexos como análise, síntese e avaliação (Bloom, 1973).

O estudo foi feito em disciplinas de Física Geral, no ciclo básico da Universidade Federal do Rio Grande do Sul (UFRGS), durante o segundo semestre de 1973. Destinou-se a responder a três questões específicas:

* Comunicação apresentada na XXVI Reunião Anual da SBPC, Recife, julho de 1974.

Trabalho parcialmente subvencionado pela CAPES.

** Do Instituto de Física da Universidade Federal do Rio Grande do Sul.

¹ Entende-se por questão de múltipla escolha o item que consiste de um enunciado seguido por um certo número de respostas, uma das quais é a certa ou a melhor e, por questão de resposta livre, o item em que o enunciado não é acompanhado pelas respostas.

1. Há diferença entre avaliar por múltipla escolha (ME) ou avaliar por resposta livre (RL) objetivos de aprendizagem definidos operacionalmente?
2. Questões que exigem níveis de comportamento mais complexos tendem a apresentar diferença maior entre os escores obtidos através de questões de ME e de RL do que questões que exigem níveis menos complexos?
3. Há correlação entre os escores obtidos de itens formulados sob a forma de ME e de RL, isto é, o teste é fidedigno?

PROCEDIMENTO EXPERIMENTAL

1. População Alvo

A experiência foi realizada com estudantes de Física Geral, Física III (FIS 104), da UFRGS. A disciplina Física III é freqüentada por estudantes de Engenharia, Física, Geologia, Matemática e Química. O programa de ensino estabelece uma carga horária de seis horas semanais, durante um semestre letivo, sendo desenvolvidos conteúdos sobre ondas, ótica, relatividade, teorias ondulatória e corpuscular da luz, noções de física do átomo e do núcleo. Bibliografia adotada: Halliday & Resnick (1973) e Beiser (1969). As aulas constaram de breve exposição do conteúdo pelos professores, seguida de trabalho com pequenos grupos de alunos. Essas aulas foram baseadas essencialmente nos objetivos de aprendizagem definidos operacionalmente, os quais foram avaliados nas questões das provas analisadas. De um total de 250 alunos, foram escolhidos aleatoriamente dois grupos de 40.

2. Fases do Desenvolvimento das Provas

Os objetivos de aprendizagem foram operacionalmente definidos envolvendo conhecimento, compreensão e aplicação dos conteúdos selecionados para serem ministrados.

Baseados nos objetivos de aprendizagem, nos comportamentos envolvidos e nos conteúdos selecionados foi elaborada uma tabela de especificação (Vianna, 1973). Essa tabela serviu para orientação na elaboração dos itens. A elaboração dos itens obedeceu a certas orientações que visavam a não debilitar uma questão na sua estruturação (Vianna, 1973 e Noll, 1965).

A revisão dos itens foi feita pelo próprio elaborador dos itens e pelos outros professores da disciplina. Foram levadas em conta, na revisão, as orientações estabelecidas para a elaboração dos itens e toda e qualquer dificuldade técnica ou de conteúdo apontada por um revisor, procurando-se evitar questões dúbias.

A montagem do teste foi feita com a preocupação de não trazer dúvidas para o aluno na forma de responder as questões. Para isso, cada prova teve uma primeira folha, de rosto, onde, além da identificação do aluno, da escola e da verificação, aparecia uma série de instruções e, quando necessário, dados necessários para responder às questões. Seguiam-se, então, os itens, metade de múltipla escolha e metade de resposta livre em cada prova, que apareceram intercalados para evitar preferência na ordem de apresentação.

Foram formuladas duas provas, com a única diferença de que as questões de múltipla escolha da primeira (prova A) foram formuladas como questões de resposta livre na segunda (prova B) e as questões de múltipla escolha da prova B foram questões de resposta livre na prova A, como ilustra o esquema abaixo:

Questão	Objetivo	Prova A	Prova B
1	1	ME	RL
2	2	RL	ME
3	3	ME	RL

Um exemplo de itens é reproduzido a seguir:

PROVA A

1. No estudo da difração de uma fenda única, qual deverá ser a largura a fim de que o segundo mínimo da luz de comprimento de onda λ ocorra na posição angular $\theta = 30^\circ$?

- A) $\lambda/4$
- B) $\lambda/2$
- C) λ
- D) 2λ
- E) 4λ

2. Os dois faróis de um automóvel que se aproxima de um observador estão separados de 1,22 m. Determinar a que distância máxima aproximada a vista humana os distinguirá um do outro? Admitir o diâmetro da pupila de $5 \cdot 10^{-3}$ m e o comprimento de onda de $5 \cdot 10^{-10}$ m.

PROVA B

1. Os dois faróis de um automóvel que se aproxima de um observador estão separados 1,22 m. A que distância máxima aproximada a vista humana os distinguirá um do outro? Admitir o diâmetro da pupila de $5 \cdot 10^{-3}$ m e o comprimento de onda de $5 \cdot 10^{-10}$ m.

- A) 5 km
- B) 8 km
- C) 10 km
- D) 15 km
- E) 25 km

2. No estudo da difração de uma fenda única, determinar a largura da fenda a fim de que o segundo mínimo da luz de comprimento de onda λ ocorra na posição angular $\theta = 30^\circ$

As provas foram misturadas para evitar que um determinado grupo ou turma recebesse apenas uma das provas, a prova A, por exemplo. Na aplicação dos testes, o professor distribuiu aleatoriamente as provas. Não foram respondidas perguntas durante a realização da prova.

A correção dos testes de resposta livre obedeceu a um critério previamente estabelecido, foi anônima e por questão. A correção dos itens de múltipla escolha foi feita na própria grade, na folha de rosto, onde o aluno assinalou a resposta escolhida em cada item. Ao item certo foi atribuído um ponto.

Os dados principais do trabalho são os acertos ou escores dos itens das provas, registrados aluno por aluno. Temos então dois grupos de dados.

ANÁLISE DOS DADOS DA EXPERIÊNCIA

Foi usado o nível de significância 0,01 para os testes estatísticos. A escolha das provas para as amostras foi aleatória.

1. Comparação dos Dados

Para a comparação das médias dos escores dos testes, foi usado o teste t (Spiegel, 1971), uma vez que a população apresentou distribuição normal.

Não há diferença significativa entre as médias dos dois tipos de testes, conforme mostram os resultados obtidos dessa análise estatística, apresentados nas Tabelas 1 e 2. Isso ocorre para todos os níveis de comportamento. A diferença tende a surgir, favorecendo a ME, para níveis mais complexos. Se os escores dos itens de múltipla escolha fossem corrigidos, levando em conta o acerto casual (Vianna, 1973), a diferença entre as médias ficaria ainda mais atenuada.

2. Fidedignidade dos Testes

Realizando as correlações entre os escores dos itens pares e os escores dos ímpares em cada prova, estimou-se a fidedignidade pela fórmula de Spearman-Brown. A fidedignidade também foi obtida pela fórmula 21 do método de Kuder-Richardson (KR_{21}) (5). Os resultados são mostrados na Tabela 3 e os coeficientes de correlação e fidedignidade foram considerados dentro do intervalo aceitável.

TABELA 1 — RELAÇÃO ENTRE TESTES DE MÚLTIPLA ESCOLHA (ME) E RESPOSTA LIVRE (RL)

Comportamento	Q_1	Q_2	N_1	N_2	\bar{X}_1	\bar{X}_2	S_1	S_2	t
Conhecimento	6	6	40	40	4,60	4,73	1,02	0,86	-0,60
Compreensão	8	8	40	40	5,02	4,55	1,56	1,72	1,27
Aplicação	11	11	40	40	8,25	7,60	1,81	1,93	1,53
Todos	25	25	40	40	18,08	16,87	2,70	3,41	1,72

Q: número de questões.

N: número de elementos da amostra

\bar{X} : média dos escores

S: desvio padrão

t: valor de t baseado no "Student's t".

índice 1: refere-se aos itens de ME da prova A

índice 2: refere-se aos itens de RL da prova B

TABELA 2 — RELAÇÃO ENTRE TESTES DE MÚLTIPLA ESCOLHA (ME) E RESPOSTA LIVRE (RL)

Comportamento	Q_3	Q_4	N_3	N_4	\bar{X}_3	\bar{X}_4	S_3	S_4	t
Conhecimento	6	6	40	40	4,77	4,57	0,96	1,26	0,79
Compreensão	8	8	40	40	5,38	4,98	1,61	1,46	1,15
Aplicação	11	11	40	40	8,83	7,93	1,36	1,90	2,40
Todos	25	25	40	40	18,98	17,48	3,18	3,41	2,01

índice 3: refere-se aos itens de ME da prova B

índice 4: refere-se aos itens de RL da prova A

TABELA 3: FIDEDIGNIDADE DOS TESTES

Prova	r	r_{sb}	KR_{21}
A	0,69	0,82	0,72
B	0,50	0,67	0,72

r: coeficiente de correlação

r_{sb} : coeficiente de fidedignidade de Spearman-Brown

KR_{21} : coeficiente de fidedignidade de Kuder-Richardson

CONCLUSÕES

Com base nos dados do estudo podemos concluir:

1. Ao comparar a média dos escores de testes de resposta livre com a média dos escores de testes de múltipla escolha em Física Geral, verifica-se que não ocorre diferença significativa. Isso ocorre para todos os itens e para os itens dos níveis de conhecimento, compreensão e avaliação separadamente.

2. A diferença de médias tende a aumentar, a favor da ME, embora não tenha sido significativa ao nível de 1%, à medida que níveis de comportamento mais complexos são exigidos. Essa tendência é razoável considerando que a percentagem de acertos das questões mais complexas é menor do que a das mais fáceis, o que propicia um maior acerto casual para as questões de ME.
3. Os coeficientes de correlação entre as partes de ME e de RL das provas são aceitáveis, sugerindo que as duas partes tem elementos em comum. Os coeficientes de fidedignidade obtidos levam-nos a concluir que podemos ter confiança no instrumento de medida utilizado.
4. Tanto os testes de ME como os de RL são bons instrumentos de medida de diferentes níveis de comportamento em Física Geral, realçada a importância do planejamento de um teste, estabelecendo não apenas o que medir, mas também como medir.

Todas as conclusões acima estão sujeitas às usuais limitações de erros de amostragem e incertezas experimentais.

REFERÊNCIAS BIBLIOGRÁFICAS

- BLOCK, J. H. 1971. *Mastery Learning — Theory and Practice*. Holt, Rinehart and Winston, Inc., New York.
- BLOOM B. S. et al. 1973. *Taxionomia de Cbjetivos Educacionais; Dominio Cognitivo*. Editora Globo, Porto Alegre.
- HALLIDAY, D. & RESNICK R. 1973. *Física*. Ao Livro Técnico S. A., Rio de Janeiro.
- BEISER, A. 1969. *Conceitos de Física Moderna*. Polígono e Editora da Universidade de São Paulo, São Paulo.
- VIANNA, H. M. 1973. *Testes em Educação*. IBRASA e Fundação Carlos Chagas, São Paulo.
- NOLL, V. H. 1965. *Introdução às Medidas Educacionais*. Livraria Pioneira Editora, São Paulo.
- SPIEGEL, M. R. 1961. *Theory and Problems of Statistics*. Schaum Publishing Co., New York.

[Recebido para publicação em junho de 1975]